Establishing intra- and inter-vendor reproducibility of T1 relaxation time measurements with 3T MRI

Lee, Yoojin ; Callaghan, Martina F ; Acosta-Cabronero, Julio ; Lutti, Antoine ; Nagy, Zoltan

Abstract: PURPOSE: Parametric imaging methods (e.g., T1 relaxation time mapping) have been shown to be more reproducible across time and vendors than weighted (e.g., T1-weighted) images. The purpose of this work was to more extensively evaluate the validity of this assertion. METHODS: Seven volunteers underwent twice-repeated acquisitions of variable flip-angle T1 mapping, including B1 + calibration, on a 3T Philips Achieva and 3T Siemens Trio scanner. Intra-scanner and inter-vendor T1 variability were calculated. To determine T1 reproducibility levels in longitudinal settings, or after changing hardware or software, four additional data sets were acquired from two of the participants; one participant was scanned on a different 3T Siemens Trio scanner and another on the same 3T Philips Achieva scanner but after a software upgrade. RESULTS: Intra-scanner variability of voxel-wise T1 values was consistent between the two vendors, averaging 0.7/0.7/1.3/1.4% in white matter/cortical gray matter/subcortical gray matter/cerebellum, respectively. We observed, however, a systematic bias between the two vendors of https://doi.org/10.0/7.8/8.6/10.0%, respectively. The T1 bias across two scanners of the same model was greater than intra-scanner variability, although still only at 1.4/1.0/1.9/2.3%, respectively. A greater bias was identified for data sets acquired before/after software upgrade in white matter/cortical gray matter (3.6/2.7%) whereas variability in subcortical gray matter/cerebellum was comparable (1.7/1.9%). CONCLUSION: We established intra- and inter-vendor reproducibility levels for a widely used T1 mapping protocol. We anticipate that these results will guide the design of multi-center studies, particularly those encompassing multiple vendors. Furthermore, this baseline level of reproducibility should be established or surpassed during the piloting phase of such studies.

DOI: https://doi.org/10.1002/mrm.27421

Establishing Intra- and Inter-vendor Reproducibility of T_1 Relaxation Time Measurements with 3T MRI

Yoojin Lee[1], Martina F. Callaghan[2], Julio Acosta-Cabronero[2], Antoine Lutti[3] and Zoltan Nagy[1]

[1] Laboratory for Social and Neural Systems Research, University of Zurich, Switzerland

[2] Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, London, UK

[3] LREN, Department for Clinical Neurosciences, CHUV, Lausanne, Switzerland

Address Correspondence to Yoojin Lee
Laboratory for Social and Neural Systems Research
University Hospital Zurich
Rämistrasse 100
P.O. Box 149
Zürich
Switzerland

Running title: Multi-Vendor Reproducibility of T_1 Mapping

25 Pages
5 Figures
2 Tables
4,680 words

Key words: Reproducibility, T_1 relaxation, multi-vendor, 3T, parametric imaging, bias

ABSTRACT

Purpose: Parametric imaging methods – e.g. T_1 relaxation time mapping – have been shown to be more reproducible across time and vendors than weighted (e.g. T_1-weighted) images. The purpose of this work was to more extensively evaluate the validity of this assertion.

Methods: Seven volunteers underwent twice-repeated acquisitions of variable flip-angle T_1 mapping, including B_1^+ calibration, on a 3T Philips Achieva and 3T Siemens Trio scanner. Intra-scanner and inter-vendor T_1 variability were calculated. To determine T_1 reproducibility levels in longitudinal settings, or after changing hardware or software, four additional datasets were acquired from two of the participants; one participant was scanned on a different 3T Siemens Trio scanner and another on the same 3T Philips Achieva scanner but after a software upgrade.

Results: Intra-scanner variability of voxel-wise T_1 values was consistent between the two vendors, averaging 0.7/0.7/1.3/1.4% in WM/cortical GM/subcortical GM/cerebellum, respectively. We observed, however, a systematic bias between the two vendors of 10.0/7.8/8.6/10.0%, respectively. The T_1 bias across two scanners of the same model was greater than intra-scanner variability, although still only at 1.4/1.0/1.9/2.3% respectively. A greater bias was identified for datasets acquired before/after software upgrade in WM/cortical GM (3.6/2.7%) while variability in subcortical GM/cerebellum was comparable (1.7/1.9%).

Conclusion: We established intra- and inter-vendor reproducibility levels for a widely used T_1 mapping protocol. We anticipate that these results will guide the design of multi-center studies, particularly those encompassing multiple vendors. Furthermore, this baseline level of reproducibility should be established or surpassed during the piloting phase of such studies.

Key words: Reproducibility, T_1 relaxation, multi-vendor, 3T, parametric imaging, bias

INTRODUCTION

Magnetic resonance imaging (MRI) is ubiquitous in both clinical diagnostics and basic research alike, and there is an increasing demand for quantitative (qMRI) methods and imaging biomarkers in general (1-8). Apart from basic scientific research into the properties of matter, measuring relaxation times (9), magnetization transfer (MT) (10), diffusion constants (11) and other quantitative parameters is of interest because together they can help characterize the underlying micro-structural organization of the tissue. A particular application of qMRI methods is the construction of invertible biophysical models to access histological metrics directly from MRI data (5) in vivo.

A further motivation behind the use of qMRI methods is the potential to obtain measures with a high degree of reproducibility across time points, scanner manufacturers, scanner models and sites. This is particularly attractive for multi-center studies (12). However, in contrast to this expectation, multiple studies have reported differences in data quality and analysis results from repeated scans (13-18), pointing out an immediate need for developing methods to investigate and quantify reproducibility (2,19-22), improving scanner design and optimizing qMRI acquisition methods.

In an effort to quantify reproducibility, this paper investigates the intra- and inter-vendor reproducibility of an established T_1 relaxation time mapping protocol. Although, many such protocols exist (23) dating back to the invention of NMR (24,25), we will focus on the variable flip angle (VFA) method (26-28). This method relies on appropriately spoiled (9,29) gradient echo (SPGR) images (30) acquired at different flip angles, together with an accompanying radio-frequency transmit field (B_1^+) map (31,32) to account for B_1^+ inhomogeneities that would otherwise lead to quantification errors, particularly at higher magnetic field strengths. We chose this framework because it is widely used (33-36), extensively optimized (37-42) and because a reliable analytical framework exists (43) for the estimation of the voxel-wise T_1 relaxation from the collected data. Few such investigations involving human volunteers and multiple sites (12,44) have been conducted. In particular, none exists in the literature that investigates T_1 mapping at 3T across multiple sites, vendors and time points. Our work demonstrates how to run a pre-study to establish reproducibility levels, which could be used in the design of future multi-center trials.

METHODS

Data Acquisition

Seven healthy volunteers (4 males, age = 34.3 ± 7.3 yrs) were scanned twice on two scanners from different vendors located at different scanning sites with maximum eight weeks apart between acquisitions at two sites. All scanning was performed with local ethics approval at each institute and a signed written informed consent was obtained before scanning. One scanner was a 3T Siemens Trio (henceforth referred to as S1; Siemens Healthcare, Erlangen, Germany) equipped with a 32-channel head coil, the other was a 3T Philips Achieva scanner (henceforth referred to as P1; Philips Healthcare, Best, The Netherlands) equipped with an 8-channel head coil. One of the seven volunteers was additionally scanned on another 3T Siemens Trio scanner (S2) with an identical 32-channel head coil and installed at the same site as S1 with 23 days apart from the test-retest scan mentioned above. The two scanners S1 and S2 were installed next to each other, a few years apart and tuned to slightly different resonance frequencies to avoid cross talk between. Otherwise they were configured with the same hardware and software (VB17A). Another one of the participants was scanned on the same 3T Philips Achieva scanner but after a software upgrade from Release 3.2.3 to 5.1.7 ($P1_u$) after five months from the test-retest scan.

The protocol to estimate T_1 relaxation time includes two 3D multi-echo SPGR images with different flip angles (6° and 21° resulting in PD- and T_1-weighted images, respectively) (43). Data were acquired with 0.8 mm isotropic resolution, TR = 25 ms, TE1/ΔTE = 2.34/2.3 ms, non-selective RF excitation, and RF spoiling phase increment = 137° for all scanners. To achieve two different flip angles, only the duration of the RF pulse was changed while the amplitude remained the same to minimize RF non-linearity effects (39). Two spoiler gradients were used; one right after the readout in the readout direction and the other right before the next RF excitation in the partition direction. The spoiler gradient moments in the readout direction were determined to achieve at least 6π dephasing across a voxel. The spoiler gradient moments in partition direction was 31.20 mT·ms/m for all scanners. Differences in the SPGR sequences between the Siemens and Philips scanners are summarized in Table 1.

For the correction of RF transmit field inhomogeneities, a B_1^+ map was additionally estimated, either from a 3D echo planar imaging (EPI) acquisition of spin-echo (SE) and stimulated-echo (STE) with 11 different refocusing flip angles (45) (Siemens, i.e. S1 and S2) or from the actual flip angle imaging (AFI) method (46) (Philips, i.e. P1 and $P1_u$). Imaging

parameters for the SE/STE-based B_1^+ mapping method were: TR = 500 ms, TE for SE/STE = 37.1 ms, two-fold generalized autocalibrating partial parallel acquisition (GRAPPA) (47) in each phase-encoded direction, and scan time = 3.0 min. The refocusing flip angles were decreased from 230° to 130° in steps of 10°. An additional dataset of B_0 field map was acquired for the correction of EPI image distortions (scan time = 2.2 min). For the AFI B_1^+ mapping method, the 3D gradient-echo images were acquired with two interleaved TRs (TR1/TR2 = 46/138 ms), TE = 2.2 ms, nominal flip angle = 60°, spoiler gradient moments for TR1/TR2 = 931.8/1971.0 mT·ms/m (48), RF spoiling phase increment = 39°, sensitivity encoding (SENSE) (49) factor = 1.7, and scan time = 5.2 min. The acquisition protocols for the two B_1^+ maps were set to ensure similar total acquisition times.

For each participant these three acquisitions (two SPGR images and a B_1^+ map) were acquired twice on the same day with the participant removed from the head coil and repositioned in between to include more potential sources of variability. This procedure was repeated on each of the S1 and P1 scanners to assess the intra-scanner and inter-vendor test/re-test variability. One of the participants was scanned four additional times on four consecutive days on the S2 scanner to establish the longitudinal reproducibility as well as the variability between two identical scanners (i.e. S1 vs. S2), i.e. intra-vendor, inter-scanner variability. The acquisition protocol on scanner S2 was identical to that of scanner S1 except the B_1^+ mapping protocol was collected first, followed by the SPGR images. Four additional datasets were acquired on four consecutive days from another participant to establish the longitudinal reproducibility on the same Philips scanner but after a software upgrade. Hence these data are designated as $P1_u$. Changing the Siemens scanner and allowing a software upgrade in the Philips scanner during the study were considered advantageous because they afforded the chance to test reproducibility in more realistic settings.

On the scanners S1 and S2, after the full calibration set that determines transmit power, shim setting, receiver gain and centre frequency was performed at the outset, no scanner calibration was performed between the three acquisitions except for the centre frequency readjustment for the B_1^+ and B_0 mapping acquisitions. Likewise, on the scanner P1 and $P1_u$, the full calibration set was performed before the first acquisition and only the centre frequency and receiver gain were readjusted for the following acquisitions. The difference in the receiver gains between three acquisitions was corrected for using the scaling information saved in the image header (14).

Data Analysis

All data analysis was performed with custom-made functions written in MATLAB (The Mathworks Inc., Natick, MA, USA) and relying on SPM12 (Wellcome Trust Centre for Neuroimaging, UCL, UK). In order to minimise the confounding effect of participant motion, two out of 38 datasets were excluded due to visible motion artifacts. The multi-echo PD- and T_1-weighted SPGR images were averaged over the five shortest TEs to increase the signal-to-noise ratio (SNR) (50). Because the transmit RF field map varies smoothly in space the estimated B_1^+ maps were low-pass filtered by a 3D Gaussian kernel with full width at half maximum of $4 \times 4 \times 4$ mm^3 to reduce the noise level. The mean T_1-weighted image and filtered B_1^+ map were aligned to the mean PD-weighted image of the same scanning session by a rigid-body transformation. For the alignment of the filtered B_1^+ map acquired with the SE/STE-based B_1^+ mapping method the sum of square image of all SE images was calculated and corrected for EPI image distortion with the B_0 map. This undistorted sum of square image was aligned to the mean PD-weighted image and the transformation matrix obtained from this alignment was applied to the filtered B_1^+ map. Similarly, the 3D gradient-echo image with TR2 was used for the alignment of AFI-based B_1^+ map. The resulting three aligned images were used to determine voxel-wise T_1 values using the equation derived from the rational approximation of the SPGR signal (43). To avoid a potential source of bias, the influence of imperfect RF spoiling was corrected using the approach proposed in Ref. (41) with the correction factors adapted to the SPGR acquisition parameters used here. To make a comparison between the estimated T_1 maps from the different acquisition sessions, rigid-body motion correction was performed for all T_1 maps from the same participant. The same-day scan/rescan T_1 maps on each of the four scanners (i.e. S1, S2, P1 and $P1_u$) were subtracted from each other and the voxel-wise difference was used to calculate a percent difference (i.e. $100 \times$ difference / mean) as a marker for intra-scanner reproducibility. A similar calculation provided the inter-scanner reproducibility measure, where the subtracted images were acquired on different scanners (e.g. between S1 and P1) or the same scanner before/after the software upgrade.

The automated anatomical labeling (AAL) (51) atlas was used to generate reproducible regions of interest (ROIs) for cortical gray matter (GM), subcortical GM, and cerebellum in Montreal neurological institute (MNI) space. Here, subcortical GM includes hippocampus, amygdala, caudate, putamen, pallidum and thalamus. These masks were then inverse-warped into each participant's native space using a deformation field obtained in the normalization of each T_1-weighted image into MNI space (52). Additionally, each R_1 map in

native space was segmented into GM and white matter (WM) using SPM12 (53). Since the R1 map is expected to have very little spatially varying bias field, very heavy bias regularization was used in the segmentation process. A conservative threshold of 0.9 was used to generate the participant-specific GM and WM masks using the output probability maps. Finally, we multiplied the native GM mask with the three inverse-normalized ROIs from the AAL atlas. Together with the native WM mask this provided four ROIs that were used for subsequent extraction of anatomically-specific summary statistics. Given that the reproducibility was not normally distributed, median and interquartile range (IQR) values were calculated as summary metrics from the voxel-wise values of the test/re-test variability images within each ROI.

RESULTS

To demonstrate the quality of the T_1 mapping scheme used in this work, one of the datasets, namely two SPGR images (averaged PD- and T_1-weighted images over the five shortest TEs) and a B_1^+ map, are shown in Fig. 1 along with the resulting T_1 map for a representative participant.

Figure 2 shows the intra-scanner and inter-vendor variability for S1 and P1 scanners from one representative participant. Intra-scanner variability maps showed high reproducibility for both scanners. While the intra-scanner variability was noisier for the S1 scanner (Fig. 2a), there was an observable level of low spatial frequency variation in the variability map of the P1 scanner (Fig. 2b). The actual spatial pattern of this low frequency variation was unique for each of the seven participants (see e.g. Fig. 5a-c). As shown in the inter-vendor variability map (Fig. 2c), there was a systematic global bias between T_1 measurements from S1 and P1 scanners, with T_1 values from S1 being lower than those from P1. A similar bias was observed for the other six participants (not shown; but see histograms in Fig. 3).

T_1 histograms for the four ROIs are shown in Fig. 3a-d for all seven participants, scanned twice both on S1 (solid line) and P1 (dashed line) scanners. All T_1 distributions from S1 were shifted towards lower T_1 values relative to those from P1, which highlights a systematic measurement bias across the two scanners. Histograms of voxel-wise T_1 variability are shown in Fig. 3e-h. While intra-scanner variability (thin solid and dashed lines) for both S1 and P1 were approximately zero-centered for all ROIs/participants, the median of the inter-vendor variability between S1 and P1 ranged from -11.5% to -4.7%.

Median and IQR values for the histograms shown in Fig. 3 are summarized in Table 2. For all ROIs and all participants the absolute median values for intra-scanner test/re-test variability were less than 3.0% for S1 and less than 3.8% for P1. Absolute median values for inter-vendor variability of all possible combinations shown in Table 2 ranged between 3.8% and 13.2% across the seven participants.

Figure 4a-c show the T_1 variability maps for the repeatedly scanned single participant (participant #6) within (Fig. 4a-b) and between (Fig. 4c) the S1 and S2 scanners. Histograms of two T_1 maps from S1 (dashed) and four T_1 maps from S2 (solid) for the four tissue segments are presented in Fig. 4d. Figure 4e shows histograms of voxel-wise T_1 variability within and between S1 and S2 scanners. Histograms of intra-scanner variability (S1 – S1 and S2 – S2) were all centered near zero, i.e. median of -0.4/-0.7/-1.8/1.2% in S1 and 0.1/0.9/1.5/0.4% in S2 for WM/cortical GM/subcortical GM/cerebellum, respectively. Medians for inter-scanner, intra-vendor variability (S1 – S2) were somewhat larger (1.4/1.0/1.9/2.3% in WM/cortical GM/subcortical GM/cerebellum, respectively).

Figure 5 summarizes results for the two datasets from P1 and the four datasets from $P1_u$ that were collected on participant #1. Figure 5a and 5b show T_1 variability maps within P1 and $P1_u$, i.e. before and after the software upgrade, respectively. An exemplar T_1 variability map between P1 and $P1_u$ is shown in Fig. 5c. Histograms for voxel-wise T_1 relaxation time and its variability maps within WM/cortical GM/subcortical GM/cerebellum are shown in Fig. 5d and 5e, respectively. While all four T_1 maps from $P1_u$ had reproducible histograms, each showed a consistent bias with respect to the T_1 maps from P1 (i.e. before the software upgrade). Histograms of voxel-wise T_1 variability for P1 and $P1_u$ were approximately zero-centered (median of -0.2/-0.3/-0.3/0.2% for P1 and 0.2/0.2/0.1/0.4% for $P1_u$ in WM/cortical GM/subcortical GM/cerebellum, respectively), indicating that four additional measurements on four consecutive days were reproducible. However, a small—but consistent—bias was observed between P1 and $P1_u$ datasets, i.e. median offset of -3.6/-2.7/-1.7/-1.9% in WM/cortical GM/subcortical GM/cerebellum, respectively.

DISCUSSION

We have quantified the level of variability for a T_1 mapping acquisition protocol by comparing data acquired repeatedly from two 3T scanner models made by two different manufacturers. The average of absolute medians for intra-scanner test/re-test variability across the seven participants and both scanners (i.e. S1 and P1) were 0.7/0.7/1.3/1.4% in

WM/cortical GM/subcortical GM/cerebellum, respectively, whereas that for inter-vendor variability was larger (10.0/7.8/8.6/10.0%) across the seven participants. We also noted that a difference could exist between two scanners of the same model or even between the same scanner with different software versions.

We observed a consistent bias in our comparison in that T_1 relaxation time estimates were 8 to 10% higher on average when calculated from data acquired on the Philips scanner compared to that from the Siemens scanner (Figs. 2 and 3). Some of this bias can be corrected by subtraction. For example, an approximately 1.4% bias in T_1 relaxation time is expected (54) due to the slightly lower main magnetic field of the Siemens scanner (2.89 versus 3.00 Tesla for the Philips scanner) and hence be corrected for. One potential source of the remaining bias could be the B_1^+ mapping protocol. It has been shown previously (31) that the 3D EPI SE/STE method produces ~5% higher B_1^+ estimates than the AFI method. Considering that the estimated T_1 relaxation time is inversely proportional to the square of B_1^+ estimates, this difference in the B_1^+ estimates would make the T_1 estimates with the 3D EPI SE/STE method ~10.6% lower compared to the T_1 estimates with the AFI method. This reasoning is in line with our experimental results. Full implementation and testing of both B_1^+ mapping methods on both scanners is beyond the scope of the present article, but such work is considered an important aspect for future efforts toward investigating or eliminating this bias.

The B_1^+ maps were smoothed by a 3D Gaussian kernel to reduce the noise level. This could be problematic at brain edges due to the nearby signal loss outside of the GM. To ensure that this issue does not cause any difference in our main results, we applied edge-preserving smoothing to the B_1^+ map and compared it to the B_1^+ map with the simple smoothing. Only ~3%/~1% of the GM voxels had larger than 2% absolute difference between these two B_1^+ maps for the datasets from S1/P1. In addition, the T_1 maps calculated with these two B_1^+ maps had almost the same histograms of T_1 variability (data not shown).

Overall, the T_1 maps from S1 were noisier than those from P1 (see Fig. 2). This is likely due to the differences in acquisition schemes whereby parallel imaging with a factor of 4 was used on S1, whereas only a factor 2 was used on P1. The different filtering processes for the alleviation of the ringing artifacts and/or different implementations of unfolding algorithms by the two scanners might also introduce further differences in relative SNR during the closed-source image reconstruction steps. These factors would appear to outweigh the expected benefit of having acquired the data with a 32-channel coil on S1, as compared to only an 8-channel coil on P1.

Despite similar test/re-test intra-scanner variability distributions across sites (Fig. 3), we observed greater low-frequency spatial variability in the reproducibility of data collected on the Philips scanner (Fig. 2b). A large number of potential sources exist for such a spatial pattern, e.g. shimming performance, RF transmit voltage, center frequency and other differences in calibration settings determined by the system at the outset of scanning. Testing each of these factors in controlled experiments is beyond the scope of this work. In any event, such factors are often not under the control of the experimenter (e.g. limited information on shimming procedure is available from scanner manufacturers) and therefore will contribute to the variability in typical imaging scenarios.

Although care was taken to match acquisition parameters on the two scanners, some differences remained (Table 1).

a) The native parallel imaging method was SENSE (49) on the Philips platform while GRAPPA (47) was used on the Siemens platform.

b) Siemens data were acquired with a 32-channel head coil whereas Philips data were received with an 8-channel head coil. The different number of channels could affect the data quality especially when parallel imaging is used. The acceleration factor on the Siemens platform was also twice as high as on the Philips platform. It is also worth noting that to maintain a comparable scan time, the Philips platform, having only an 8ch coil, used a partial Fourier acquisition in one of the phase encoding directions as opposed to the two-by-two GRAPPA acceleration on Siemens.

c) Philips scanners offer a product AFI B_1^+ mapping sequence (46), which is not natively available on Siemens scanners; thus, a previously validated, in-house B_1^+ mapping method (31,35,36,45) was utilized. As previously discussed, discrepancies between the two B_1^+ mapping approaches might be responsible for the observed systematic bias between the two scanners.

d) The exact shimming procedure performed on each scanner is largely unknown to the end-user. It is therefore possible that they are different for each scanner type. Even if the procedure was identical, the quality of the shim may be variable due to the fact that the magnet as well as the static shimming implementation as well as the linear and higher order shimming gradient designs differ between the two manufacturers.

e) Siemens Trio scanners utilize a tube amplifier for RF transmission while Philips scanners are equipped with a solid-state variant. The inherent stability and linearity can be expected to differ between both design types.

It is possible that the low-frequency spatial variability on test/re-test measures for the Philips scanner (compare for example Fig. 2a and Fig 2b) or the consistent bias across the two scanners (Fig 2c and Table 2) stem from some combination of the differences listed above. By recording additional information at scan time, it may be possible to implement correction schemes for some of these. For example, collecting k-space data and reconstructing offline would enable the use of a common unfolding algorithm across data types, the B_1^+ mapping procedure could also be harmonized, or the same head coil could be used for the two different scanners. Any resulting improvement in reproducibility would need to be evaluated. Indeed it may be true that if the above differences in the acquisition procedures were eliminated the bias between the two manufacturers would also disappear.

It is also important to note that with scanner-specific shimming procedures, RF optimization and hardware design, most users—including expert ones—might need to concede that some aspects of this T_1 mapping protocol will not be possible to fully harmonize. This has important consequences for multi-center studies, which are often motivated by the premise that several sites can recruit the required number of individuals more efficiently for a given study. Here, we show how the results presented above may be used to optimize the design of multi-center studies, by considering an example ROI in the precentral gyrus (i.e. primary motor cortex). Based on the variability across the participants (i.e. standard deviation of 32-44 ms), the required sample size would be 3-5 for either the S1 or the P1 scanner to identify a 5% difference in the mean between two groups (a significance level of 0.05 and a power of 0.80). If both S1 and P1 were used in conjunction to form a multi-center setting, a sample size of 9-19 would be required to identify the same difference in the mean T_1 relaxation time between two groups (the standard deviation being 58-84 ms). This substantial increase in the required number of participants is due to the often-ignored but yet significant variance originating from the inherent difference that may occur between measurements made on scanners at different sites. In the present case, this additional variance was dominated by a systematic bias between two scanners, but it may equally be a random factor in case of other scanners. In cases where a consistent systematic bias is detected, it could be partially eliminated before analyzing datasets, by subtracting/adding the bias averaged over participants. In the above example, if the systematic bias of -5.4% measured inside precentral gyrus was removed, the standard deviation would go down to 29-50 ms, resulting in the required sample size of 2-7. Please note however that due to the dependence between the bias in B_1^+ and the true T_1 relaxation time this correction cannot be complete.

In WM and cortical GM the impact of the software upgrade on P1 was higher than the intra-scanner or the intra-vendor/inter-scanner (i.e. S1 vs. S2) variabilities although it was lower than the inter-vendor variability. However, in subcortical GM and the cerebellum the bias introduced by a software upgrade on the P1 scanner was comparable to variability introduced by changing from S1 to S2. Given that no hardware was changed and that the same sequence settings were used before and after the upgrade, these differences likely stem from the image reconstruction or the preparation steps before the acquisition (e.g. the way B_1 is calibrated, the way optimal shim values are found, etc), which are not accessible to the end-users. This bias highlights the importance of software configurations of the scanner. For example, image reconstructions in multi-center study could be harmonized by using open-source, vendor-independent platforms (55) to alleviate such an inter-vendor bias.

With the data collected it was possible to estimate also the relative proton density (rPD) maps (i.e. we normalized all voxels so that the mean rPD inside WM would be 69%). These maps were used to assess reproducibility analogously to that for T_1 relaxation time and investigate whether the biases found in these two measures are correlated. Please see Supporting Information for the details.

Reproducibility has been assessed by many others in the past (13-18). While these reports are significant in raising awareness of the issues on reproducibility, none of them focused in T_1 mapping specifically. Reproducibility of T_1 mapping was investigated across several vendors but at 1.5T (44) or at 3T but on identical scanners installed at different sites (12). Making direct comparisons between these previous findings and our results is difficult however for several reasons. E.g. at 1.5T transmit field inhomogeneity is a much more benign issue and therefore Deoni et al. (44) proceed without a B_1^+ mapping sequence in their protocol. Nevertheless, the intra-site coefficient of variation they reported in Table 4 was higher for all of their seven participants than the percent differences we found in any of the four ROIs (Table 2). This is particularly notable because the variability of B_1^+ maps is a significant contributor to variability in T_1 maps (48). Had they collected a B_1^+ map it is possible that their coefficient of variance would have been even higher. An exhaustive comparison to the study of Weiskopf et al. (12) is also difficult because their study design was significantly different (i.e. used a single scanner model and did not scan repeatedly on the same scanner). They reported that the inter-site bias was lower than 3.1%, which is comparable to the observation we made on a single subject (i.e. 1.4/1.0/1.9/2.3% for WM/cortical GM/subcortical GM/cerebellum respectively and please see Fig. 4) that was scanned twice on two Siemens scanners (i.e. S1 and S2).

It may be considered a limitation of this study that other manufacturers (e.g. GE) were not involved. Indeed, in light of the present results, establishing both intra- and inter-vendor reproducibility levels for some or all other scanner models would be extremely beneficial. However, including one or more scanner models would not change the main results of the present study - that the reproducibility of (multi-center) studies that involve scanners from different manufacturers could suffer from significant biases. Apart from involving only two scanners, the sample size is also small and therefore our results are not generalizable. Nevertheless, our aim was simply to point out that reproducibility across scanner types may be worse than the reproducibility on any given scanner. Therefore, we recommend conducting a pre-study on the specific scanners involved, before embarking on a full-scale investigation involving multiple scanner types (note: both hardware and software differences can be relevant). Please note that the longitudinal scans and hence the biases we reported between S1 and S2 as well as between P1 and $P1_u$ were performed on a single subject. Therefore, statistical significance for these biases cannot be established and should be taken here as an indication only.

In conclusion, the present study reported on test/re-test reproducibility measures for a widely used T_1 mapping protocol, which on average returned approximately 1% intra-scanner, 1% intra-vendor/inter-scanner and 8-10% inter-vendor variability across the whole brain. While already indicating a high level of reproducibility for the intra-scanner and intra-vendor/inter-scanner scenarios, future developments in acquisition and data processing will likely result in further improvements for both intra-scanner and, more particularly, inter-vendor reproducibility. At present these results represent the most comprehensive study of T_1 mapping variance across 3T scanner types using a variable flip-angle protocol. It is recommended that multi-center studies consider such level of variance in both study design and result interpretation. Our demonstration of a consistent bias of approximately 3% between T_1 values measured before and after a scanner software upgrade indicates the importance of not only considering hardware stability but also software configurations. Therefore, even when using a single scanner, care must be taken to ensure system consistency, both in terms of hardware and software, for the duration of each study. Further work across the community, encompassing both researchers and manufacturers, is needed to more extensively investigate and ultimately enhance reproducibility.

ACKNOWLEDGMENTS

YL and ZN were supported by the Swiss National Science Foundation (grant number: 31003A_166118). MFC is supported by the MRC and Spinal Research Charity through the ERA-NET Neuron joint call (MR/R000050/1). The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome Trust [203147/Z/16/Z]. AL was supported by the Roger De Spoelberch Foundation. The authors thank Dr. Roger Luechinger for sharing his expertise of the Philips MRI platform and Mr. Karl Treiber for scanning support.

REFERENCES

1. Tofts P. Quantitative MRI of the brain: Measuring changes caused by disease. John Wiley & Sons; 2005.
2. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res 2015, Feb;24(1):9-26.
3. Prescott JW. Quantitative imaging biomarkers: The application of advanced image processing and analysis to clinical and preclinical decision making. J Digit Imaging 2013, Feb;26(1):97-108.
4. Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Abramson RG, Burton KR, et al. Clinical utility of quantitative imaging. Acad Radiol 2015, Jan;22(1):33-49.
5. Weiskopf N, Mohammadi S, Lutti A, Callaghan MF. Advances in mri-based computational neuroanatomy: From morphometry to in-vivo histology. Curr Opin Neurol 2015, Aug;28(4):313-22.
6. Cheng HL, Stikov N, Ghugre NR, Wright GA. Practical medical applications of quantitative MR relaxometry. J Magn Reson Imaging 2012, Oct;36(4):805-24.
7. Deoni SC. Quantitative relaxometry of the brain. Top Magn Reson Imaging 2010, Apr;21(2):101-13.
8. Mezer A, Yeatman JD, Stikov N, Kay KN, Cho NJ, Dougherty RF, et al. Quantifying the local tissue volume and composition in individual brains with magnetic resonance imaging. Nat Med 2013, Dec;19(12):1667-72.
9. Haacke EM. Magnetic resonance imaging: Physical principles and sequence design. New York: J. Wiley & Sons; 1999.

10. Henkelman RM, Stanisz GJ, Graham SJ. Magnetization transfer in MRI: A review. NMR Biomed 2001, Apr;14(2):57-64.
11. Le Bihan D. Diffusion and perfusion magnetic resonance imaging applications to functional MRI. New York: Raven Press; 1995.
12. Weiskopf N, Suckling J, Williams G, Correia MM, Inkster B, Tait R, et al. Quantitative multi-parameter mapping of R1, PD(*), MT, and R2(*) at 3T: A multi-center validation. Front Neurosci 2013;7:95.
13. Focke NK, Helms G, Kaspar S, Diederich C, Tóth V, Dechent P, et al. Multi-site voxel-based morphometry--not quite there yet. Neuroimage 2011, Jun 1;56(3):1164-70.
14. Chenevert TL, Malyarenko DI, Newitt D, Li X, Jayatilake M, Tudorica A, et al. Errors in quantitative image analysis due to platform-dependent image scaling. Translational Oncology 2014;7(1):65-71.
15. Malyarenko DI, Newitt D, J Wilmes L, Tudorica A, Helmer KG, Arlinghaus LR, et al. Demonstration of nonlinearity bias in the measurement of the apparent diffusion coefficient in multicenter trials. Magn Reson Med 2015, May 2.
16. Jakab A, Werner B, Piccirelli M, Kovács K, Martin E, Thornton JS, et al. Feasibility of diffusion tractography for the reconstruction of intra-thalamic and cerebello-thalamic targets for functional neurosurgery: A multi-vendor pilot study in four subjects. Front Neuroanat 2016;10:76.
17. Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, et al. Reliability of mri-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. Neuroimage 2006, Aug 1;32(1):180-94.
18. Bauer CM, Jara H, Killiany R, Alzheimer's Disease Neuroimaging Initiative. Whole brain quantitative T2 MRI across multiple scanners with dual echo FSE: Applications to AD, MCI, and normal aging. Neuroimage 2010, Aug 15;52(2):508-14.
19. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology standards for quantitative imaging biomarkers. Radiology 2015, Dec;277(3):813-25.
20. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. Stat Methods Med Res 2015, Feb;24(1):27-67.
21. Obuchowski NA, Reeves AP, Huang EP, Wang XF, Buckler AJ, Kim HJ, et al. Quantitative imaging biomarkers: A review of statistical methods for computer algorithm comparisons. Stat Methods Med Res 2015, Feb;24(1):68-106.

22. Abramson RG, Burton KR, Yu JP, Scalzetti EM, Yankeelov TE, Rosenkrantz AB, et al. Methods and challenges in quantitative imaging biomarker development. Acad Radiol 2015, Jan;22(1):25-32.
23. Kingsley PB. Methods of measuring spin-lattice (T1) relaxation times: An annotated bibliography. Concepts in Magnetic Resonance 1999;11(4):243-76.
24. Drain LE. A direct method of measuring nuclear spin-lattice relaxation times. Proceedings of the Physical Society. Section A 1949;62(5):301.
25. Hahn EL. An accurate nuclear magnetic resonance method for measuring spin-lattice relaxation times. Physical Review 1949;76(1):145.
26. Christensen KA, Grant DM, Schulman EM, Walling C. Optimal determination of relaxation times of fourier transform nuclear magnetic resonance. Determination of spin-lattice relaxation times in chemically polarized species. The Journal of Physical Chemistry 1974;78(19):1971-7.
27. Fram EK, Herfkens RJ, Johnson GA, Glover GH, Karis JP, Shimakawa A, et al. Rapid calculation of T1 using variable flip angle gradient refocused imaging. Magn Reson Imaging 1987;5(3):201-8.
28. Gupta RK. A new look at the method of variable nutation angle for the measurement of spin-lattice relaxation times using FT NMR. J Magn Reson 1977;25:231-.
29. Zur Y, Wood ML, Neuringer LJ. Spoiling of transverse magnetization in steady-state sequences. Magn Reson Med 1991, Oct;21(2):251-63.
30. Haase A, Frahm J, Matthaei D, Hanicke W, Merboldt KD. FLASH imaging. Rapid NMR imaging using low flip-angle pulses. Journal of Magnetic Resonance 1986;67(2):258-66.
31. Lutti A, Hutton C, Finsterbusch J, Helms G, Weiskopf N. Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T. Magn Reson Med 2010, Jul;64(1):229-38.
32. Pohmann R, Scheffler K. A theoretical and experimental comparison of different techniques for B1 mapping at very high fields. NMR Biomed 2013, Mar;26(3):265-75.
33. Deoni SC, Peters TM, Rutt BK. High-resolution T1 and T2 mapping of the brain in a clinically acceptable time with DESPOT1 and DESPOT2. Magn Reson Med 2005, Jan;53(1):237-41.
34. Draganski B, Ashburner J, Hutton C, Kherif F, Frackowiak RS, Helms G, Weiskopf N. Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). Neuroimage 2011, Apr 15;55(4):1423-34.

35. Sereno MI, Lutti A, Weiskopf N, Dick F. Mapping the human cortical surface by combining quantitative T(1) with retinotopy. Cereb Cortex 2013, Sep;23(9):2261-8.
36. Lutti A, Dick F, Sereno MI, Weiskopf N. Using high-resolution quantitative mapping of R1 as an index of cortical myelination. Neuroimage 2013, Jun 10;93 Pt 2:176-88.
37. Deoni SC, Peters TM, Rutt BK. Determination of optimal angles for variable nutation proton magnetic spin-lattice, T1, and spin-spin, T2, relaxation times measurement. Magn Reson Med 2004, Jan;51(1):194-9.
38. Deoni SC. High-resolution T1 mapping of the brain at 3T with driven equilibrium single pulse observation of T1 with high-speed incorporation of RF field inhomogeneities (DESPOT1-HIFI). J Magn Reson Imaging 2007, Oct;26(4):1106-11.
39. Lutti A, Weiskopf N. Optimizing the accuracy of T1 mapping accounting for RF non-linearities and spoiling characteristics in FLASH imaging. In Proceedings of the 21st Annual Meeting of ISMRM, Salt Lake City, USA, 2013. Abstract 2478.
40. Cheng HL, Wright GA. Rapid high-resolution T(1) mapping by variable flip angles: Accurate and precise measurements in the presence of radiofrequency field inhomogeneity. Magn Reson Med 2006, Mar;55(3):566-74.
41. Preibisch C, Deichmann R. Influence of RF spoiling on the stability and accuracy of T1 mapping based on spoiled FLASH with varying flip angles. Magn Reson Med 2009, Jan;61(1):125-35.
42. Wood TC. Improved formulas for the two optimum VFA flip-angles. Magn Reson Med 2015, Jul 10;74(1):1-3.
43. Helms G, Dathe H, Dechent P. Quantitative FLASH MRI at 3T using a rational approximation of the ernst equation. Magn Reson Med 2008, Mar;59(3):667-72.
44. Deoni SC, Williams SC, Jezzard P, Suckling J, Murphy DG, Jones DK. Standardized structural magnetic resonance imaging in multicentre studies using quantitative T1 and T2 imaging at 1.5 T. Neuroimage 2008, Apr 1;40(2):662-71.
45. Lutti A, Stadler J, Josephs O, Windischberger C, Speck O, Bernarding J, et al. Robust and fast whole brain mapping of the RF transmit field B1 at 7T. PLoS One 2012;7(3):e32379.
46. Yarnykh VL. Actual flip-angle imaging in the pulsed steady state: A method for rapid three-dimensional mapping of the transmitted radiofrequency field. Magn Reson Med 2007, Jan;57(1):192-200.
47. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 2002, Jun;47(6):1202-10.

48. Lee Y, Callaghan MF, Nagy Z. Analysis of the precision of variable flip angle t1mapping with emphasis on the noise propagated from RF transmit field maps. Front Neurosci 2017;11:106.
49. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: Sensitivity encoding for fast MRI. Magn Reson Med 1999, Nov;42(5):952-62.
50. Helms G, Dechent P. Increased SNR and reduced distortions by averaging multiple gradient echo signals in 3D FLASH imaging of the human brain at 3T. J Magn Reson Imaging 2009, Jan;29(1):198-204.
51. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 2002, Jan;15(1):273-89.
52. Callaghan MF, Helms G, Lutti A, Mohammadi S, Weiskopf N. A general linear relaxometry model of R1 using imaging data. Magn Reson Med 2015, Mar;73(3):1309-14.
53. Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005, Jul 1;26(3):839-51.
54. Rooney WD, Johnson G, Li X, Cohen ER, Kim SG, Ugurbil K, Springer CS. Magnetic field and tissue dependencies of human brain longitudinal 1H2O relaxation in vivo. Magn Reson Med 2007, Feb;57(2):308-18.
55. Hansen MS, Sørensen TS. Gadgetron: An open source framework for medical image reconstruction. Magn Reson Med 2013, Jun;69(6):1768-76.

FIGURE LEGEND

Figure 1. A representative example of SPGR images with two different flip angles (i.e. averaged PD- and T_1-weighted images over the five shortest TEs), a B_1^+ map and the resulting T_1 map.

Figure 2. (a-b) Intra-scanner test/re-test variability of T_1 maps acquired on **(a)** S1 and **(b)** P1 scanners. **(c)** Inter-vendor variability of T_1 maps between S1 and P1 scanners. Four different inter-vendor variability maps can be calculated from the two datasets acquired on each scanner. One of these is shown here. The other three maps are not displayed but Table 2 summarizes those results separately for each combination.

Figure 3. Histograms of voxel-wise test/re-test T_1 estimates **(a-d)** acquired on the S1 and P1 scanners and their voxel-wise variability **(e-h)** inside each of four ROIs (i.e. WM **(a/e)**,

cortical GM **(b/f)**, subcortical GM **(c/g)**, and cerebellum **(d/h)**) for all seven participants. Different colors represent different participants (abbreviated as Parti in figure legend). In total, there are 2 (S1 and P1) x 2 (number of repetitions) x 7 (number of participants) = 28 curves in **(a-d)** and 3 (two intra-scanner and one inter-scanner variabilities) x 7 (number of participants) = 21 curves in **(e-h)**. Note: the *x* and *y* axes are scaled differently for each subplot and each histogram contains the results of all possible combinations of the four datasets acquired on the S1 and P1 scanners.

Figure 4. The results from four longitudinal datasets on the S2 scanner and comparison to the two datasets on the S1 scanner (different participant from Fig. 2). **(a)** Intra-scanner variability map for the S1 scanner. **(b)** Intra-scanner variability between the first two maps from the longitudinal datasets acquired on the S2 scanner. **(c)** Intra-vendor, inter-scanner variability between S1 and S2. **(d)** Histograms of the voxel-wise two T_1 estimates from two acquisitions on S1 (dashed) and four acquisitions on S2 (solid), inside the WM (black), cortical GM (blue), subcortical GM (red) and cerebellum (cyan). There are 6 (two repetitions on S1 and four repetitions on S2) x 4 (four ROIs) = 24 curves in total. **(e)** Histograms of the voxel-wise intra-scanner T_1 variability on the S1 (thin solid lines) and S2 (dashed lines), as well as the voxel-wise intra-vendor, inter-scanner T_1 variability between S1 and S2 (thick solid lines). There are 3 (two intra-scanner and one intra-vendor, inter-scanner variability) x 4 (four ROIs) = 12 curves in total. Note: The histograms of S1 – S2 contain the eight different intra-vendor, inter-scanner T_1 variability maps calculated from two T_1 maps on S1 and four T_1 maps on S2.

Figure 5. The results from four longitudinal datasets on $P1_u$ and comparison to the two datasets on P1 for participant #1. **(a)** Intra-scanner variability map for the P1 scanner. **(b)** Intra-scanner variability map from the first two T_1 maps acquired on the $P1_u$ scanner. **(c)** Intra-scanner, inter-software version T_1 variability between P1 and $P1_u$. **(d)** Histograms of two T_1 maps from P1 (dashed lines) and four T_1 maps from $P1_u$ (solid lines), for the WM (black), cortical GM (blue), subcortical GM (red) and cerebellum (cyan) segments (24 curves in total). **(e)** Histograms of intra-scanner T_1 variability on the scanner P1 (thin solid lines) and $P1_u$ (dashed lines), and intra-scanner, inter-software version T_1 variability between P1 and $P1_u$ (thick solid lines) for the four ROIs. Eight different intra-vendor, inter-software version T_1 variability maps were calculated from two T_1 maps on P1 and four T_1 maps on $P1_u$ and combined for the histograms of P1 – $P1_u$ (12 curves in total).

SUPPORTING INFORMATION

Supporting Figure S1. Intra-scanner **(a,b)** and inter-vendor variability **(c)** of rPD maps for participant #1.

Supporting Figure S2. The histograms of rPD **(a-d)** and rPD variability **(e-h)** inside four ROIs for all seven participants. Note that all histograms for rPD variability were centered near 0, which means that intra-scanner and inter-vendor reproducibility were high for rPD maps.

Supporting Figure S3. The relationship between T_1 variability and rPD variability for intra-scanner (S1 – S1 and P1 – P1) and inter-vendor (S1 – P1) settings.

TABLES

Table 1. Comparison of acquisition parameters for SPGR images

	Siemens	Philips
# of slices	224 sagittal partitions	204 sagittal partitions
# of echoes	8	5
Spoiler gradient in readout direction	88.10 mT/m·ms	93.34 mT/m·ms
Parallel imaging	Four-fold GRAPPA (2×2)	Two-fold SENSE
Half Fourier	none	0.625 in partition direction
RF excitation pulse	sinc with 2 zero-crossings and TBW[a] = 6	sinc with 1 zero-crossing and TBW[a] = 4.3
Readout bandwidth	488 Hz/pixel	497.1 Hz/pixel
Total scan time	7.1 min	7.6 min

[a]TBW: time-bandwidth product

Table 2. Median ± IQR of voxel-wise test/re-test T_1 variability [%]

		Intra-scanner variability [%]		Inter-vendor variability [%]			
		$S1^{1st} - S1^{2nd}$	$P1^{1st} - P1^{2nd}$	$S1^{1st} - P1^{1st}$	$S1^{1st} - P1^{2nd}$	$S1^{2nd} - P1^{1st}$	$S1^{2nd} - P1^{2nd}$
Participant #1	WM	-0.43 ± 12.30	-0.21 ± 9.54	-10.90 ± 12.30	-11.07 ± 12.52	-10.53 ± 11.65	-10.72 ± 11.68
	Cortical GM	0.21 ± 12.14	-0.18 ± 10.13	-7.95 ± 13.38	-8.10 ± 13.15	-8.18 ± 13.30	-8.28 ± 12.87
	Subcor GM[a]	2.32 ± 16.90	-0.30 ± 13.09	-8.09 ± 15.21	-8.30 ± 16.13	-10.46 ± 15.09	-10.69 ± 15.06
	Cerebellum	0.67 ± 17.97	0.26 ± 13.82	-11.12 ± 19.36	-10.84 ± 19.41	-12.05 ± 18.52	-11.64 ± 18.27
Participant #2	WM	-0.28 ± 10.59	0.11 ± 8.08	-6.93 ± 10.48	-6.82 ± 10.53	-6.66 ± 10.11	-6.56 ± 9.99
	Cortical GM	0.28 ± 10.31	-0.49 ± 8.79	-4.32 ± 11.74	-4.77 ± 11.80	-4.60 ± 11.40	-5.06 ± 11.45
	Subcor GM[a]	2.04 ± 14.93	0.81 ± 10.31	-4.67 ± 13.51	-3.79 ± 13.80	-6.77 ± 13.41	-5.91 ± 13.50
	Cerebellum	-0.63 ± 16.16	-0.79 ± 12.00	-4.58 ± 16.71	-5.37 ± 16.14	-4.03 ± 17.56	-4.83 ± 17.03
Participant #3	WM	-2.33 ± 11.07	0.74 ± 8.98	-12.44 ± 11.11	-11.68 ± 10.91	-10.12 ± 10.57	-9.38 ± 10.19
	Cortical GM	-1.24 ± 11.71	0.85 ± 9.69	-9.35 ± 12.88	-8.42 ± 12.87	-7.99 ± 11.89	-7.12 ± 11.43
	Subcor GM[a]	0.40 ± 14.95	1.11 ± 12.99	-9.77 ± 14.66	-8.61 ± 14.82	-10.07 ± 14.01	-8.99 ± 13.72
	Cerebellum	0.74 ± 20.91	1.25 ± 12.31	-11.21 ± 18.73	-9.95 ± 19.20	-12.05 ± 18.29	-10.71 ± 17.79
Participant #4	WM	0.06 ± 11.98	0.11 ± 9.71	-8.95 ± 13.26	-8.91 ± 13.45	-9.02 ± 12.04	-8.95 ± 12.25
	Cortical GM	0.13 ± 13.99	-0.05 ± 10.61	-8.19 ± 15.27	-8.31 ± 15.59	-8.18 ± 14.08	-8.28 ± 14.27
	Subcor GM[a]	1.97 ± 15.99	0.84 ± 13.37	-7.68 ± 15.66	-6.71 ± 15.82	-9.37 ± 15.87	-8.63 ± 15.78
	Cerebellum	1.79 ± 21.43	2.06 ± 13.27	-11.44 ± 22.15	-9.60 ± 22.47	-13.35 ± 21.58	-11.49 ± 21.74
Participant #5	WM	1.18 ± 12.34	-0.63 ± 9.52	-10.24 ± 12.36	-10.89 ± 12.05	-11.52 ± 11.74	-12.12 ± 11.96
	Cortical GM	1.27 ± 12.4	0.52 ± 10.07	-8.12 ± 13.85	-7.52 ± 13.65	-9.44 ± 13.43	-8.83 ± 13.78
	Subcor GM[a]	2.78 ± 17.51	-0.15 ± 12.2	-9.76 ± 16.79	-9.81 ± 15.81	-12.75 ± 15.29	-12.75 ± 15.91
	Cerebellum	2.96 ± 17.68	-0.80 ± 16.70	-9.34 ± 20.81	-10.00 ± 18.24	-12.66 ± 20.72	-13.18 ± 19.29
Participant #6	WM	-1.31 ± 11.73	2.35 ± 11.06	-11.89 ± 12.25	-9.55 ± 11.62	-10.60 ± 12.91	-8.32 ± 12.22
	Cortical GM	-0.65 ± 13.48	2.23 ± 13.15	-9.63 ± 14.34	-7.33 ± 14.45	-9.00 ± 15.53	-6.76 ± 15.64
	Subcor GM[a]	-0.75 ± 17.34	1.50 ± 14.42	-9.68 ± 15.99	-8.23 ± 16.08	-8.76 ± 17.80	-7.46 ± 17.49
	Cerebellum	1.66 ± 20.80	3.79 ± 16.40	-11.29 ± 19.24	-7.39 ± 18.48	-13.20 ± 21.14	-9.23 ± 21.21
Participant #7	WM	-0.17 ± 11.61	0.26 ± 10.89	-11.47 ± 11.90	-11.37 ± 12.39	-11.35 ± 11.69	-11.19 ± 11.81
	Cortical GM	0.61 ± 11.26	1.45 ± 11.80	-9.34 ± 13.05	-7.89 ± 13.86	-9.88 ± 12.93	-8.52 ± 13.46
	Subcor GM[a]	1.22 ± 16.39	-1.96 ± 13.62	-6.51 ± 15.80	-8.46 ± 16.49	-7.81 ± 15.33	-9.74 ± 15.33
	Cerebellum	-0.29 ± 17.27	2.49 ± 15.68	-11.35 ± 19.37	-9.11 ± 19.26	-11.22 ± 19.21	-8.95 ± 18.73

[a]Subcor GM: subcortical gray matter

FIGURES

Figure 1

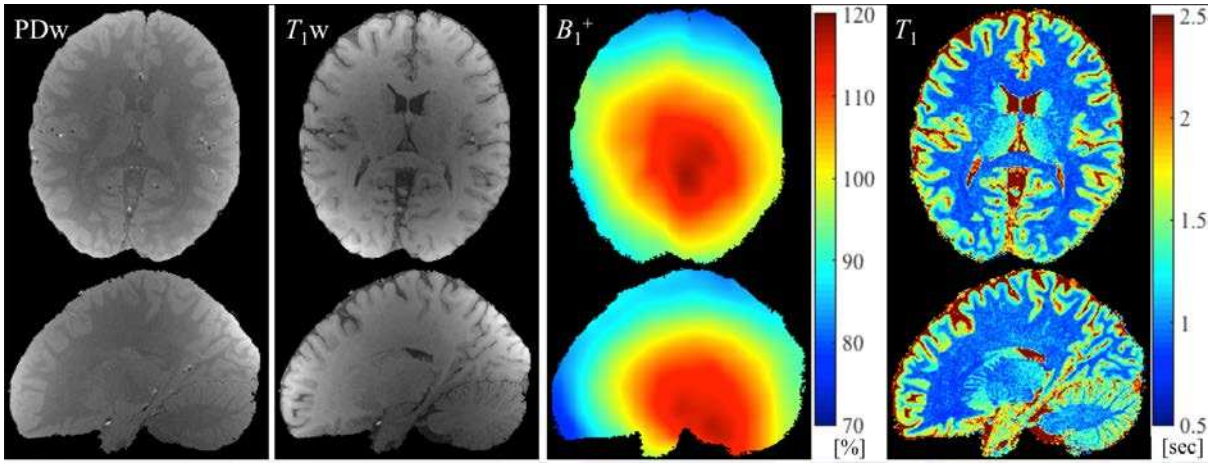


Figure 2

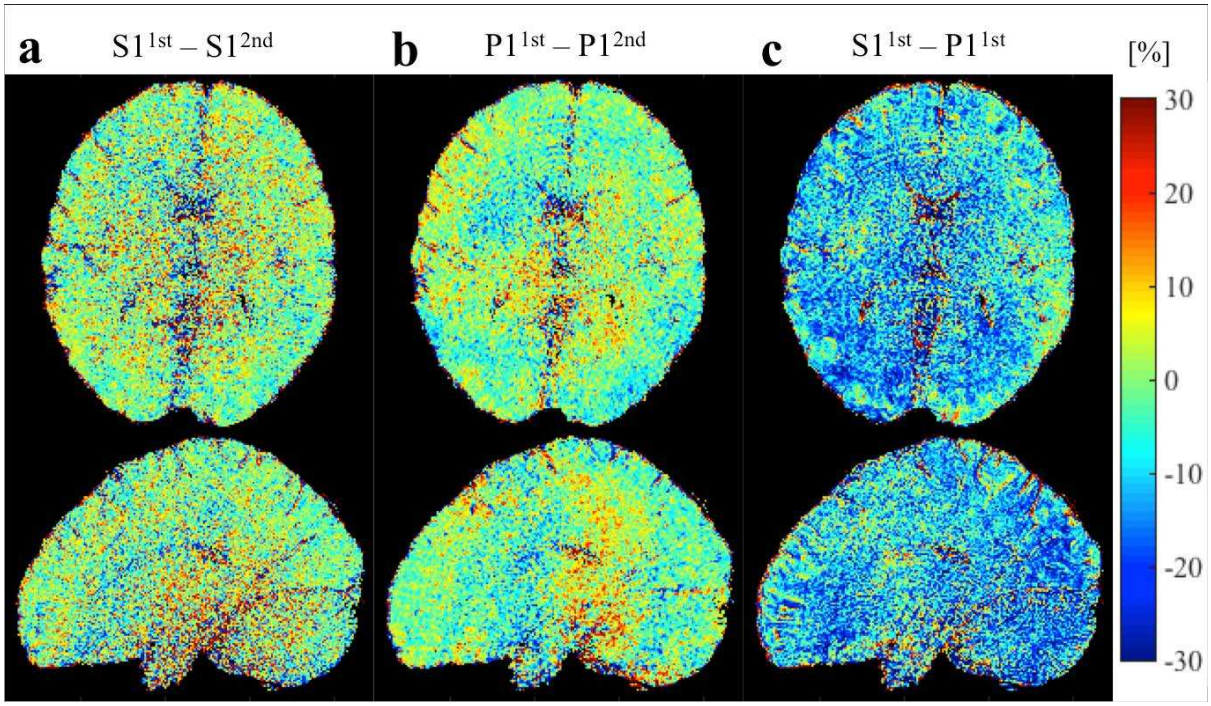


Figure 3

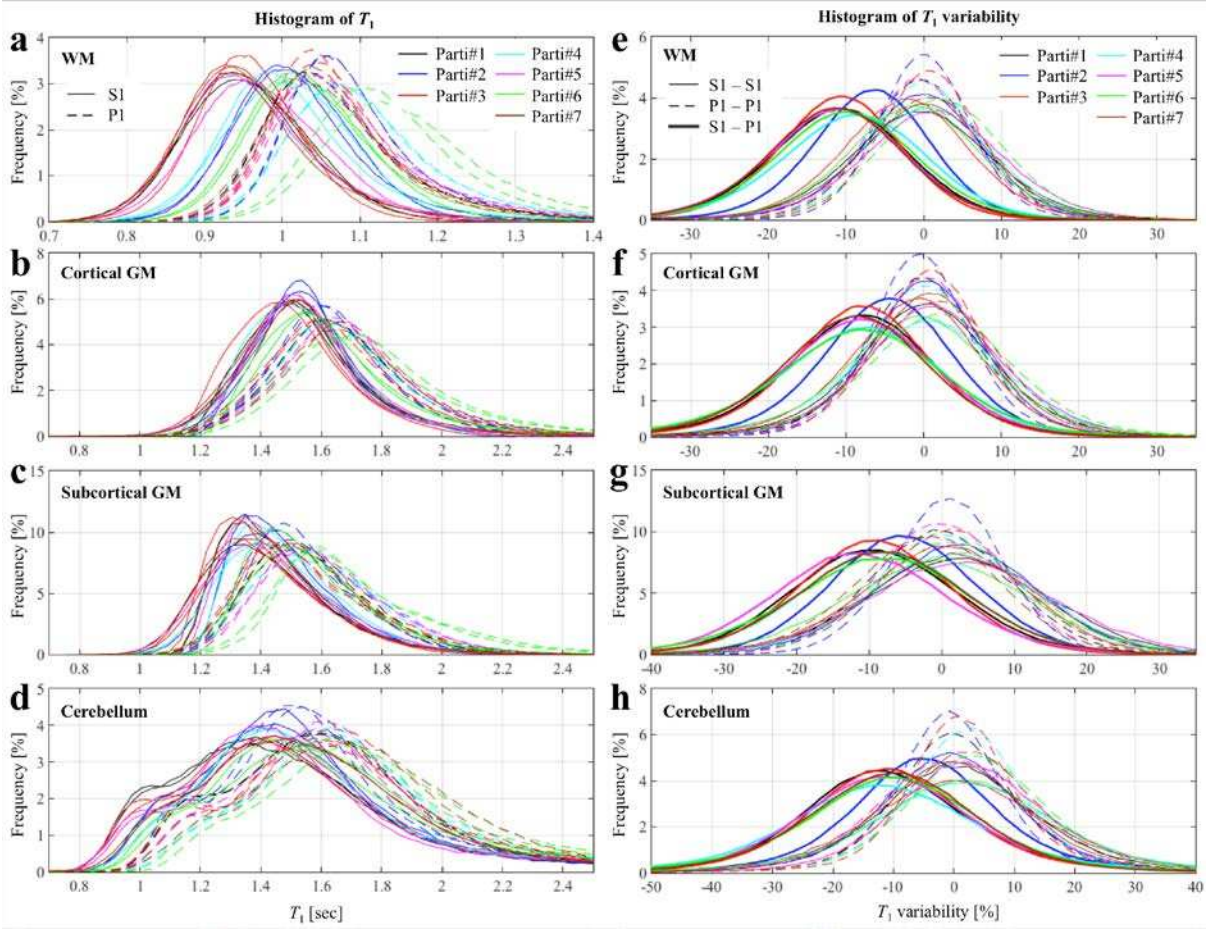

Histogram of T_1
Histogram of T_1 variability
a
WM
S1
P1
Parti#1
Parti#2
Parti#3
Parti#4
Parti#5
Parti#6
Parti#7
Frequency [%]
b
Cortical GM
c
Subcortical GM
d
Cerebellum
T_1 [sec]
e
WM
S1 – S1
P1 – P1
S1 – P1
f
Cortical GM
g
Subcortical GM
h
Cerebellum
T_1 variability [%]

Figure 4

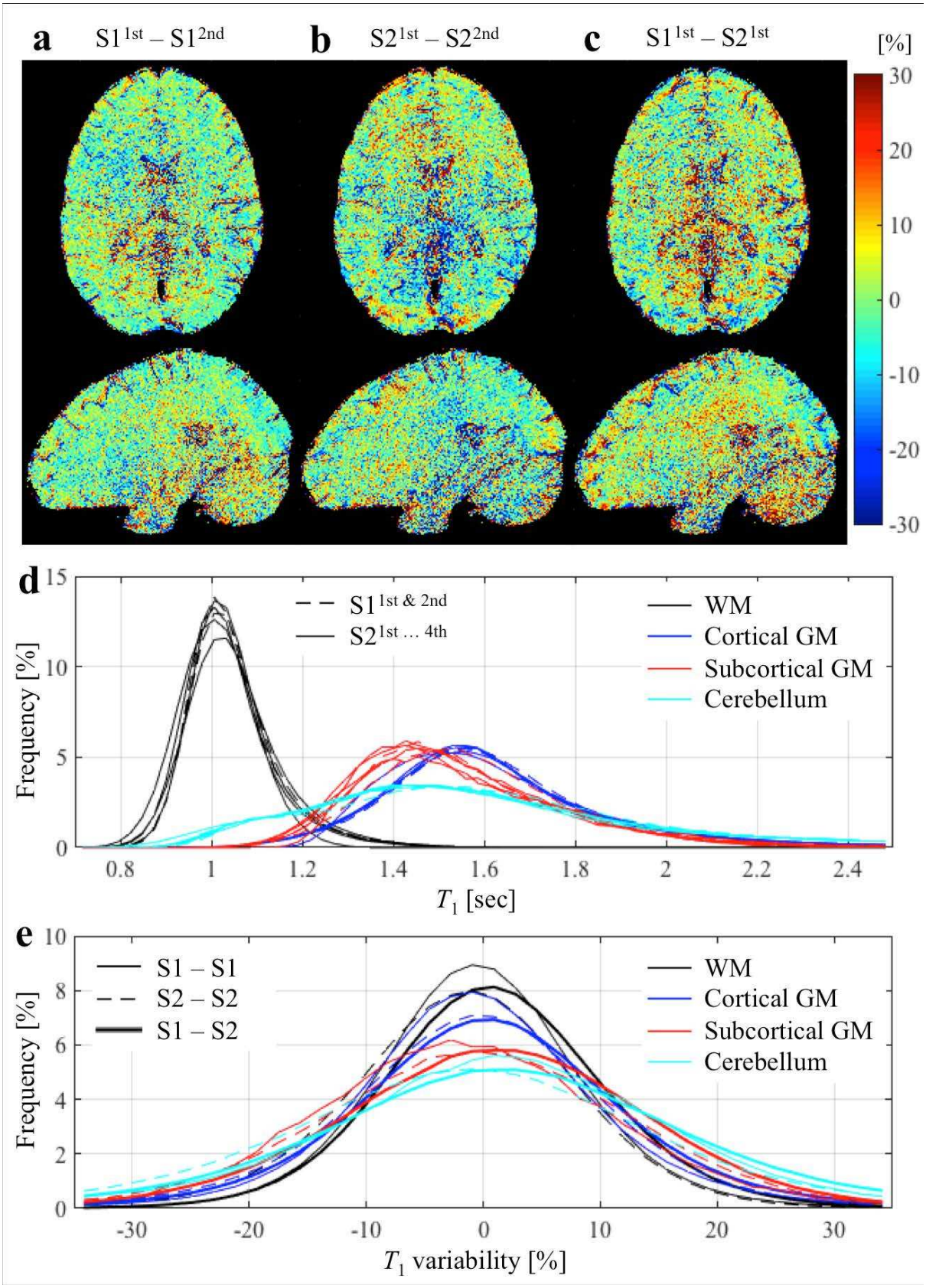
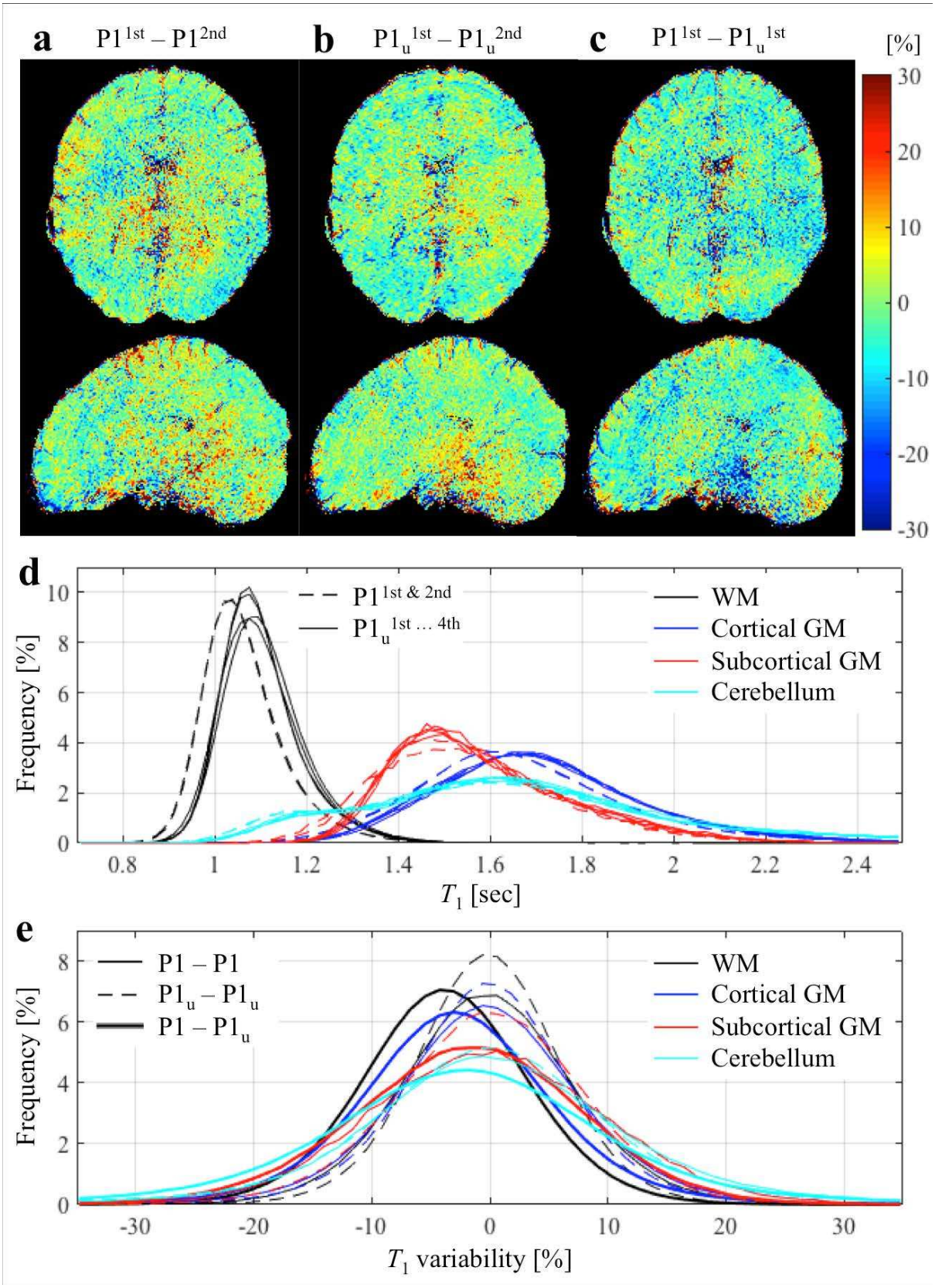

a
S1 1st – S1 2nd
b
S2 1st – S2 2nd
c
S1 1st – S2 1st
[%]
30
20
10
0
-10
-20
-30
d
Frequency [%]
S1 1st & 2nd
S2 1st … 4th
WM
Cortical GM
Subcortical GM
Cerebellum
T1 [sec]
e
Frequency [%]
S1 – S1
S2 – S2
S1 – S2
WM
Cortical GM
Subcortical GM
Cerebellum
T1 variability [%]

Figure 5


a
P1 1st – P1 2nd
b
P1u 1st – P1u 2nd
c
P1 1st – P1u 1st
[%]
30
20
10
0
-10
-20
-30
d
P1 1st & 2nd
P1u 1st … 4th
WM
Cortical GM
Subcortical GM
Cerebellum
Frequency [%]
T1 [sec]
e
P1 – P1
P1u – P1u
P1 – P1u
WM
Cortical GM
Subcortical GM
Cerebellum
Frequency [%]
T1 variability [%]